



Implementing a Storage Architecture for SQL Server using the Lenovo ThinkSystem DM7100F Storage Array

Last update: 17 June 2021

Version 1.6

Highlights benefits of ThinkSystem DM7100F SAN configurations

Presents a use case for Lenovo ThinkSystem DM7100F Storage Array

Provides deployment information and best practices

Includes a detailed bill of materials for ordering

David West
Vinay Kulkarni



Table of Contents

1	Introduction	1
2	Business value	2
3	Architectural overview	3
4	Benefits of SAN storage solutions	4
5	Deployment Prerequisites	5
6	NVMe configuration	6
6.1	Storage Virtual Machines	6
6.2	Namespaces	6
6.3	Host NQNs.....	6
6.4	NVMe Subsystem	7
6.5	FC switch zoning.....	7
6.6	Windows MPIO	8
6.7	Performance settings	8
7	Overview of SQL Server storage and high availability	10
7.1	SQL Server storage best practices	10
7.2	SQL Server high availability	11
8	Appendix: Bill of Materials	13
8.1	Storage BOM	13
	Conclusion	14
	Resources	15

1 Introduction

This document describes the configuration and best practices for running Microsoft SQL Server 2019 workloads on the ThinkSystem DM7100F storage solution from Lenovo. The DM7100F provides a robust storage solution for mission critical SQL Server workloads. Additional features included in the ONTAP software adds powerful flexibility, connectivity, and management options. An end-to-end NVMe solution is now available with the Lenovo ThinkSystem DM7100F all-flash storage and DB620S switches with Emulex LPe35000 series 32Gb Fibre Channel host adapters.

The intended audience is IT professionals, technical architects, sales engineers, and consultants to assist in planning, designing, and implementing this solution.

This document provides an overview of the business problem and business value that is addressed by having a highly available SQL Server 2019 instance on the Lenovo DM7100F storage system.

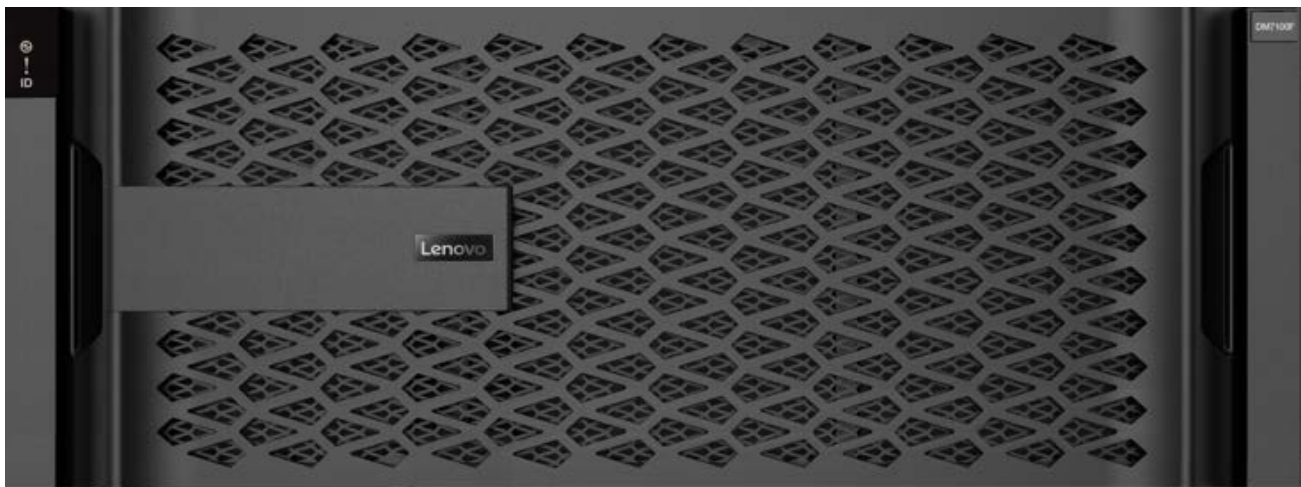


Figure 1 *Lenovo ThinkSystem DM7100F*

Lenovo ThinkSystem DM7100F is a scalable, unified, all flash storage system that is designed to provide high performance, simplicity, capacity, security, and high availability for large enterprises. Powered by the ONTAP software, ThinkSystem DM7100F delivers enterprise-class storage management capabilities with a wide choice of host connectivity options, flexible drive configurations, and enhanced data management features, including end-to-end NVMe support (NVMe over Fabrics and NVMe drives). The DM7100F is a perfect fit for a wide range of enterprise workloads, including big data and analytics, artificial intelligence, engineering and design, hybrid clouds, and other storage I/O-intensive applications.

2 Business value

The Lenovo ThinkSystem DM7100F storage offering for SQL Server provides enterprise customers a highly available, cost efficient, flexible platform for high-performance Microsoft SQL Servers, leveraging state-of-the-art all-flash hardware and the latest storage protocols like NVMe over Fibre Channel.

Whether running Online Transaction Processing (OLTP) workloads, or Data Warehouse and BI, to AI and advanced analytics over Big Data, you benefit from the resiliency the DM7100F offers. This is especially important for mission critical databases.

Additionally, since the storage supports simultaneous SCSI FC and NVMe host connections, it enables simplified migration of data to the newer NVMe storage.

Fast flash technologies allow for lower latencies, quicker response times and smarter data management in real-time for database or virtualization workloads.

3 Architectural overview

The Lenovo ThinkSystem DM7100F is a high-performance unified all-flash storage solution well suited for SQL Server workloads. The following section takes a high-level look at the architecture of the system, and the logical mapping of the storage components.

The foundation of the DM7100F is the clustered active-active dual controller, providing balanced performance and redundancy. The included web based ONTAP storage management software and GUI provides all the features and configuration options. The storage protocols supported include traditional Fibre Channel, NVMe over FC, Network Attached Storage (NAS) and high-speed iSCSI connections - depending on the configuration that is ordered. For this paper, since testing was done on a system configured for FC and NVMe/FC, the focus is on those protocols. By using the storage pool concept, the physical disks are grouped together providing a virtualized high-performance pool from which volumes are created.

The configuration follows a common SAN pattern of creating volumes from pools and assigning them to hosts. There are important differences in how these tasks are accomplished depending on the storage protocol being used. Fibre Channel uses traditional LUN concepts, while NVME/FC uses namespaces. These differences are explained in the later sections.

The volume to host mapping architecture is shown below. Section 5 in the document covers these in more detail.

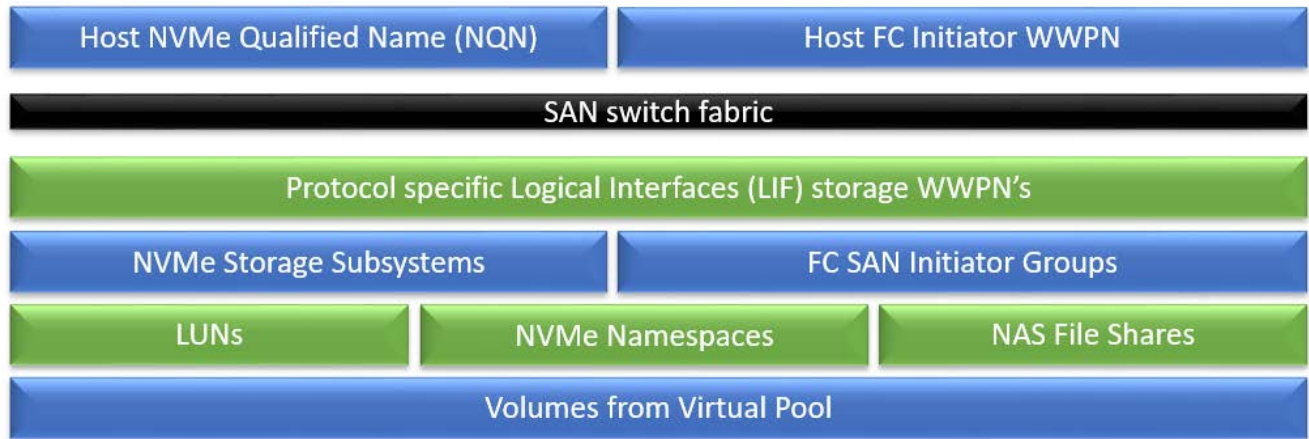


Figure 2 Overview of volume to host architecture

4 Benefits of SAN storage solutions

SANs are still widely used today, despite the trend for local storage, especially in enterprise environments where redundant, reliable, and high-performance storage is expected. In the case of the DM7100F, a single system contains two controllers, with half of the volumes owned by each of the controllers. In the event of a controller failure, all volumes temporarily fail over ownership to the other controller. Additionally, RAID volumes are normally configured to provide redundancy in case a drive fails. More complex availability models include multiple SAN systems, replicating data within and to other data centers for disaster recovery scenarios.

Some additional points for SAN vs. local or NAS solutions

- Improved security by consistent encryption management, separate data location and network
- Highest levels of up time possible due to the redundant, robust, highly available designs
- Advanced features such as deduplication, compression, replication, snap shots, data tiering, and seamless migration
- Centralized management and monitoring of data for many servers, is more efficient than managing multiple single servers with local storage
- Nearly unlimited scale that can seamlessly grow and migrate as the business requires, without impacting hosts or applications
- Better support for virtualized and clustered environments, allowing seamless migration of virtual machines to different hosts
- Centralized management of data retention, policies, and departments across an organization
- SAN performance isn't impacted by LAN or local disk bandwidth bottlenecks, its on a dedicated high-speed network
- Build on existing investments, by adding the latest technologies to existing SAN infrastructure
- Centralized backups, with only one backup server needed to manage and support high volumes of storage.
- SAN based backups are faster than network based, and free up LAN bandwidth
- More efficient disk utilization, as volumes are provisioned from disk pools and virtualized
- Support for disk-less servers and boot from SAN, providing cost savings at the server level

5 Deployment Prerequisites

There are a few items to be aware of before deploying the system. A quick start guide is included in the box that provides everything needed to deploy and the easy to follow steps.

The system needs to have redundant ethernet management, controller, and expansion interconnects attached before initializing it. The controller interconnects are direct connected and require detailed attention to the specific cables as shown in the quick start guide. The rest of the connections are to the Fibre Channel (FC) switch, as are the host connections. Be aware that the hosts cannot directly connect to the storage system.

The FC switch zones can be configured after the storage and hosts are all connected, and the WWPN port numbers are determined. NVMe uses the same zoning concepts as traditional SCSI FC, allowing both types of volumes to be mapped to hosts over the same fabric. There are separate storage target WWPNs, associated with each protocol specific Logical Interfaces, which is covered in the next section.

6 NVMe configuration

This section provides detailed guidance for the volume and host configurations for a better understanding of NVMe over Fibre Channel concepts. The topics to be covered include:

- Storage virtual machines
- Namespaces
- Host NQN's
- NVMe Subsystem (host) definition
- FC switch zoning
- Windows MPIO
- Performance settings

6.1 Storage Virtual Machines

Storage VMs (SVM) are created by the system for each storage protocol that is enabled. So, in our case there was one for NVMe and one for FC. The system usually recognizes the protocol being used and automatically assigns any resources to the correct SVM. There may be some configuration interfaces that prompt for which SVM to use, so it's important to understand the concept.

6.2 Namespaces

NVMe namespaces are synonymous with LUNs in traditional FC SAN systems. It is a logical definition that is associated with a physical volume. Namespaces use NVMe Qualified Names (NQN) from the host, in a similar manner as initiator WWPNs on a SAN.

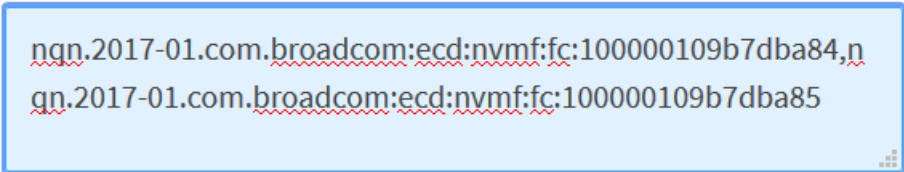
6.3 Host NQNs

When creating Namespaces, you must provide the hosts NQN strings. For the Brocade HBAs, we determined the NQN of the initiator ports by using the following formula provided by Brocade:

nqn.2017-01.com.broadcom:ecd:nvmf:fc:<factory WWPN> NOTE: Do not include colons (:) when specifying the WWPNs.

Notice that the last part of the string is the WWPN of each HBA port. This is showing both NQN's, comma separated, no space between comma.

HOST NQN



```
nqn.2017-01.com.broadcom:ecd:nvmf:fc:100000109b7dba84,n  
qn.2017-01.com.broadcom:ecd:nvmf:fc:100000109b7dba85
```

Figure 3 Sample of host NQN list when creating Namespaces

6.4 NVMe Subsystem

The NVMe Subsystem is how NVMe/FC defines hosts, like an initiator group in traditional SAN terminology. This also requires the hosts NQN strings. Below is a view of an NVMe Subsystem for the host named SQLnode2 and associated NQN's, namespaces, SVM's and volume usage.

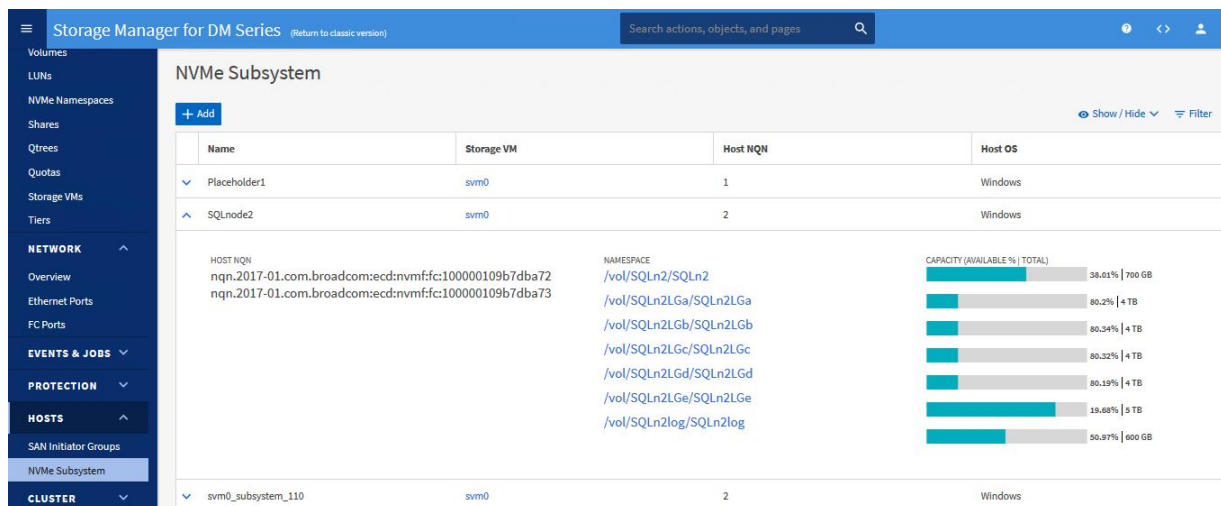


Figure 4 View of NVMe Subsystem details

6.5 FC switch zoning

The zoning for NVMe uses the same concepts as standard FC SAN. Use single initiator 1:1 zones and ensure the correct WWPN is used for the port ID's. The Logical Interface (LIF) WWPN's can be found under the Network section, Overview as shown below. The LIF WWPNs for each protocol need to be zoned.

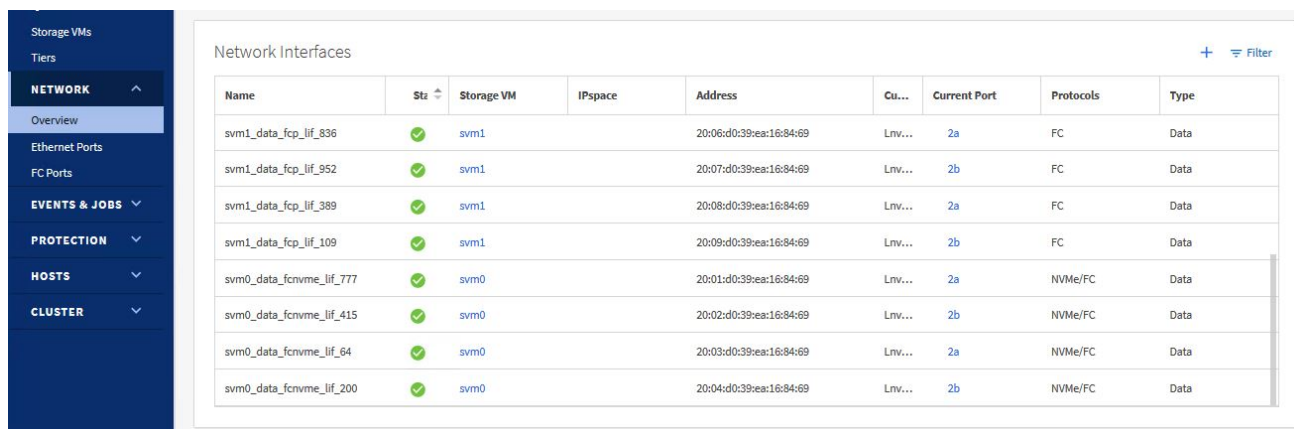


Figure 5 Where to find the LIF WWPNs used for zoning hosts to the storage

6.6 Windows MPIO

We used the native MPIO feature included with Windows for multipathing control. On the second tab of the MPIO applet it should discover the NVMe NetApp ONTAP device. Select it and click **Add** as shown below. Wait for it and reboot the server after it finishes.

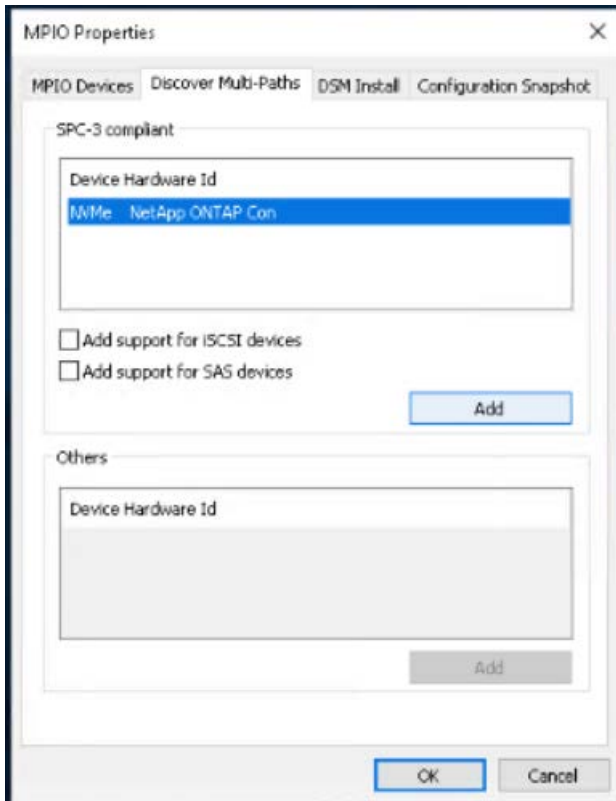


Figure 6 How to add NVMe support in Windows native MPIO configurations

6.7 Performance settings

By default, the QOS policy has a limitation set on it that limits the performance needed for enterprise workloads like SQL Server. The 10,000 IOPS limit needs to be set to unlimited, by following the steps below.

Go to Storage VMs and select the SVM used for NVMe. Scroll down and find the QOS box. Click the small arrow icon at the top right of the QOS box to open it. Select the listed SVM and chose Edit. Enter 0 (zero) in each of the boxes, as shown below.

Note that this is not a global setting, it is specific for each SVM and as a result is specific for each protocol. It is likely the FC SVM also needs the same setting for higher performance.

Edit QoS Policy Group [Tips for QoS polices](#) ✕

POLICY GROUP NAME
svm0_dpo_default

GUARANTEE (IOPS)
0

LIMIT
0 MB/s

LIMIT (IOPS)
0

Share performance limits
MB/s and IOPS limits are shared by all associated storage objects.

[Cancel](#) [Update](#)

Figure 7 Setting the QoS Policy Group to all 0's for unlimited performance

In summary, the benefits of NVMe/FC include the following.

- Fibre Channel switch zoning for NVMe is identical to what is recommended for normal FC zones
- Both NVMe and SCSI volumes can be provisioned and assigned to hosts concurrently, making it easy to migrate from SCSI FC to NVMe volumes
- Enables tiering with cold data left on slower performing SCSI volumes, hot data on NVMe
- Provides interoperability with existing Fibre Channel infrastructure and IT skill sets
- Hosts use Microsoft's native MPIO multipathing
- NVMe/FC provides significantly higher IOPS over SCSI/FC volumes on the same hardware

Some differences between NVMe storage and FC provisioning includes the following:

- Uses host NQN strings vs WWPN to map volumes
- NVMe namespaces are configured and map to volumes vs LUNs for FC
- Hosts are called NVMe subsystems composed of NQN strings

7 Overview of SQL Server storage and high availability

After the storage is configured and connected, the focus can shift to setting up and optimizing the applications. This section covers SQL Server and high availability options with Lenovo servers and storage.

7.1 SQL Server storage best practices

The following recommendations are provided to ensure the best performance of SQL Server databases on the Lenovo servers and DM7100F storage.

General settings for MS SQL solutions from Lenovo

- Update to latest firmware & driver levels on servers and all components
- Configure UEFI settings for Operating mode to Maximum performance
- Configure high availability for the OS with 2-disk Raid-1 mirror
- Configure high availability for data with ONTAP RAID-DP
- Configure high availability for log files with 2-disk Raid-1 or Raid-10 with a higher even number disks
- Use multiple DB and tempdb files, spreading them evenly across all data drives for optimal performance
- Set the power plan in Windows to high performance: Control Panel > System & Security > Power Options > High Performance
- Set the Windows virtual memory file to 4096 MB initial and maximum (not system managed) and to use the C drive only. Verify there is no page file residing on any of the SQL volumes.
- Set Windows to adjust for best performance of programs: Control Panel > System & Security > Advanced System settings
- Enable lock pages in memory using Windows Group policy tool to prevent paging of data.

Specific settings for Data Warehouse workloads

If the server is dedicated to current workload, these settings provide optimal performance for SQL DW.

- Set processor affinity for SQL Server to use all the processors in the system
- Set SQL Server Maximum Server Memory to 90% of total memory available on server
- Optionally add –T834 to SQL Server Startup parameters to set the trace flag to enable large pages for SQL Server buffer pool.
- Enable parallel processing by setting Max degree of parallelism to number of CPU in system

Additional settings for OLTP workloads

In addition to the above, configure the usual SQL OLTP performance parameters, including:

- Lightweight pooling enabled
- Max worker threads set to 3000
- Priority Boost enabled
- Recovery interval set to 32767
- Disable parallel processing by setting Max degree of parallelism to 1

7.2 SQL Server high availability

SQL Server's Availability Group (AG) feature manages data replication at the SQL layer to provide database level high availability and disaster recovery. SQL is installed on the VMs, which are clustered at the VM level. Although not covered in this document, it is also possible to use SQL Failover Cluster Instances (FCI) on VMs to provide SQL application level protection along with AGs.

SQL AG's support both synchronous or asynchronous replication. Synchronous is slightly slower performing but is always in synch with no data loss at failover as each write is verified. Asynchronous maintains fast performance but results in some data loss during failover.

The SQL Enterprise edition supports up to 8 replicas (copies) per group, and an unlimited number of availability groups. Additional readable copies can be configured for access by read-only workloads, reporting, or to perform backups from them.

Failover can be performed manually or configured to trigger automatically if the primary database becomes unavailable. After a failover occurs, the secondary copy becomes the primary copy and clients are automatically redirected to the new copy as primary. The clients are redirected by using a SQL virtual device called the AG Listener, which is registered in DNS.

For SQL Server based high availability, the servers don't use shared volumes. Each node in a SQL availability group is connected to its own storage. This storage could be separate volumes on the same storage device, or on separate storage devices in either local or remote data centers for the best redundancy. SQL Server handles the synchronization and failover of data between the volumes attached to each node.

Below is an example of SQL AGs running on VMs and a Windows Cluster.

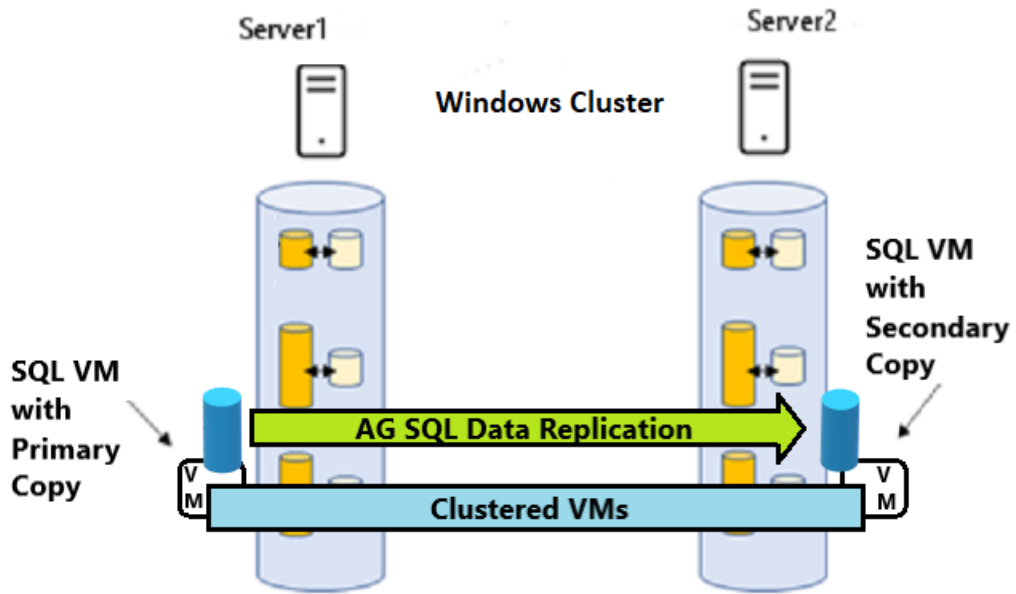


Figure 8 Example of SQL Availability Group running on VMs and a Windows failover cluster

8 Appendix: Bill of Materials

This appendix features the Bill of Materials (BOMs) for the storage system. The BOM listed is not meant to be exhaustive and must always be confirmed with the configuration tools and customer requirements.

8.1 Storage BOM

Part #	Description	Quantity
7D25CTO1WW	Controller : Lenovo ThinkSystem DM7100F All Flash Array	1
B94E	Lenovo ThinkSystem DM 4U Chassis	1
B5RJ	DM Series Premium Offering	1
B94T	Lenovo ThinkSystem DM7100 Controller	2
BAZ1	2P 100Gb PCIe Ethernet SmartIO Adapter for DM Series AFF Storage	2
B94J	Lenovo ThinkSystem DM Series 25Gb 4 Port Ethernet Mez Card	2
AV1W	Lenovo 1m Passive 25G SFP28 DAC Cable	2
AV1Z	Lenovo 1m Passive 100G QSFP28 DAC Cable	2
B4BP	Lenovo ThinkSystem Storage USB Cable, Micro-USB	1
6400	2.8m, 13A/100-250V, C13 to C14 Jumper Cord	4
7Y62CTO1WW	Storage : Lenovo ThinkSystem DM240N 2U24 NVMe Expansion Enclosure	1
B6W6	Lenovo ThinkSystem Storage NVMe 2U24 Chassis	1
B73A	ThinkSystem Storage NVMe Expansion IOM	2
BC7W	Lenovo ThinkSystem 23TB (6x 3.84TB NVMe Non-SED) Drive Pack for DM7100F	2
AV1Z	Lenovo 1m Passive 100G QSFP28 DAC Cable	4
6311	2.8m, 10A/100-250V, C13 to C14 Jumper Cord	2
6415HC1	Switch : Lenovo ThinkSystem DB620S FC SAN Switch 24x32Gb SWL SFP	1
AVG2	Lenovo DB620S FC SAN Switch (Entry)	1
AVGC	Brocade 32GB SWL SFP+ Transceiver	8
B2PB	Lenovo 3m LC-LC OM4 MMF Cable	8
6201	1.5m, 10A/100-250V, C13 to IEC 320-C14 Rack Power Cable	2

Conclusion

The Lenovo ThinkSystem DM7100F is an ideal storage platform for configuring Microsoft SQL Server to run with high availability and all flash low latency. As an end-to-end NVMe enabled storage solution, the DM7100F provides Enterprise environments flexible connectivity options, high-speed and reliable storage for their mission-critical SQL Server workloads

Resources

For more information about the topics that are described in this document, see the following resources:

- Lenovo ThinkSystem DM7001F product overview
<https://lenovopress.com/lp1271-thinksystem-dm7100f-unified-all-flash-storage-array>
- Lenovo ThinkSystem DM7001F product home page
<https://www.lenovo.com/us/en/data-center/storage/unified-storage/thinksystem-dm-series/ThinkSystem-DM-Series-All-Flash-Array/p/WMD00000375>
- Microsoft SQL Server Availability Groups documentation
<https://docs.microsoft.com/en-us/sql/database-engine/availability-groups/windows/overview-of-always-on-availability-groups-sql-server?view=sql-server-ver15>
- Microsoft SQL Server documentation
<https://docs.microsoft.com/en-us/sql/index>

Trademarks and special notices

© Copyright Lenovo 2021.

References in this document to Lenovo products or services do not imply that Lenovo intends to make them available in every country.

Lenovo, the Lenovo logo, ThinkSystem, ThinkAgile, ThinkCentre, ThinkVision, ThinkVantage, ThinkPlus and Rescue and Recovery are trademarks of Lenovo.

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both.

Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.

Intel, Intel Inside (logos), MMX, and Pentium are trademarks of Intel Corporation in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.

Information is provided "AS IS" without warranty of any kind.

All customer examples described are presented as illustrations of how those customers have used Lenovo products and the results they may have achieved. Actual environmental costs and performance characteristics may vary by customer.

Information concerning non-Lenovo products was obtained from a supplier of these products, published announcement material, or other publicly available sources and does not constitute an endorsement of such products by Lenovo. Sources for non-Lenovo list prices and performance numbers are taken from publicly available information, including vendor announcements and vendor worldwide homepages. Lenovo has not tested these products and cannot confirm the accuracy of performance, capability, or any other claims related to non-Lenovo products. Questions on the capability of non-Lenovo products should be addressed to the supplier of those products.

All statements regarding Lenovo future direction and intent are subject to change or withdrawal without notice and represent goals and objectives only. Contact your local Lenovo office or Lenovo authorized reseller for the full text of the specific Statement of Direction.

Some information addresses anticipated future capabilities. Such information is not intended as a definitive statement of a commitment to specific levels of performance, function or delivery schedules with respect to any future products. Such commitments are only made in Lenovo product announcements. The information is presented here to communicate Lenovo's current investment and development activities as a good faith effort to help with our customers' future planning.

Performance is based on measurements and projections using standard Lenovo benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput or performance improvements equivalent to the ratios stated here.

Photographs shown are of engineering prototypes. Changes may be incorporated in production models.

Any references in this information to non-Lenovo websites are provided for convenience only and do not in any manner serve as an endorsement of those websites. The materials at those websites are not part of the materials for this Lenovo product and use of those websites is at your own risk.